Appendix S1

To estimate gestational age at delivery, we developed three predictors using combinations of longitudinal maternal and cross-sectional neonatal anthropometry measures under a linear mixed modeling framework for predictors using maternal anthropometry and an added a shared parameter model to incorporate neonatal information. Theoretical details, along with simulation results and method development have been previously reported.[22] Predictors were developed in a training set, where we consider the case where we have a gold standard measurement of gestational age for all individuals. In this scenario we develop the following predictors:

1. Univariate Longitudinal Maternal Measure

   Let Y_{ij} be the jth fundal height measurement for the ith subject, and G_i be the true gold standard gestational age measurement for the ith subject. We consider the linear mixed model

   $$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1}\left(G_i + t_{ij}\right) + \varepsilon_{ij} \qquad (1)$$

   where $i = 1, \dots, N$ represents the subject, $j = 1, \dots, n_i$ represents the measurement, $\boldsymbol{b_i} = (b_{i0}, b_{i1})'$ is assumed to have a multivariate normal distribution with mean $\mathbf{0}$ and variance Σ_b, ε_{ij} is normal with mean 0 and variance σ_ε^2. This situation represents the model presented by Laird and Ware (1982).[32] We assume G_i is normally distributed with mean μ_G and variance σ_G^2 and that they do not depend on covariates. We also assume that $t_{i0} = 0$ for all i, where t_{ij} represents the time from the initial visit until the jth measurement. We can estimate the parameters using standard software, our calculations were performed using the *lme* function of the package *nlme* in R.[33] Estimations of μ_G and σ_G^2 can be obtained from the sample estimates form the training set:

   $$\hat{\mu}_G = \frac{1}{N}\sum_i G_i\,; \quad \widehat{\sigma}_G^{\,2} = \frac{1}{N-1}\sum_i (G_i - \hat{\mu}_G)^\wedge 2 \qquad (2)$$

   To predict gestational age we propose $E[G_i|\boldsymbol{Y_i}, \theta]$, where $\boldsymbol{Y_i} = (y_{i1}, \dots, y_{in_i})'$ and $\theta = \{\boldsymbol{\beta}, \Sigma_b \sigma_\varepsilon^2, \mu_G, \sigma_G^2\}$, with $\boldsymbol{\beta} = (\beta_0, \beta_1)'$.

$$E[G_i|\boldsymbol{Y_i},\theta]=\frac{\int_G\int_{\boldsymbol{b_i}}G_i\left[\prod_j f\left(y_{ij}\middle|G_i,\boldsymbol{b_i}\right)\right]g(G_i)h(\boldsymbol{b_i})dG_id\boldsymbol{b_i}}{\int_G\int_{\boldsymbol{b_i}}\left[\prod_j f\left(y_{ij}\middle|G_i,\boldsymbol{b_i}\right)\right]g(G_i)h(\boldsymbol{b_i})dG_id\boldsymbol{b_i}} \tag{3}$$

2. Multivariate Longitudinal Maternal Measure

We consider the model

$$Y_{ijk}=\beta_{0k}+\beta_{1k}t_{ij}+b_{i0k}+b_{i1k}\left(G_i+t_{ij}\right)+\varepsilon_{ijk} \tag{4}$$

where Y_{ijk} represents the jth measurement for the ith subject of the kth maternal anthropometry measure, with $i=1,\dots,N$, $j=1,\dots,n_i$ and $k=1,\dots,K$. $\boldsymbol{b_{ik}}=(b_{i01},bi_{1k},\dots,b_{i0K},b_{i1K})'$ is normally distributed with mean **0** and variance Σ_b, $\varepsilon_{ijk}\sim N(0,\sigma_\varepsilon^2)$ and $G_i\sim N(\mu_G,\sigma_G^2)$. We maintain the assumption that $t_{i0}=0$ for all i and that G_i is observed.

The estimation of (4) can be obtain by estimating the multivariate linear model when $K\leq 3$ using standard software or in the high dimensional case by fitting pairwise mixed models (Fiews and Verbeke, 2006). Estimations of μ_G and σ_G^2 are obtained from (2).

To predict gestational age we propose $E[G_i|\boldsymbol{Y_{i1}},\dots,\boldsymbol{Y_{ik}},\theta]$, where $\boldsymbol{Y_{ik}}=(Y_{i1k},\dots,y_{in_ik})'$ and $\theta=\{\boldsymbol{\beta_1},\dots\boldsymbol{\beta_k},\Sigma_b,\sigma_\varepsilon^2,\mu_G,\sigma_G^2\}$, with $\boldsymbol{\beta_k}=(\beta_{0k},\beta_{1k})'$.

$$E[G_i|\boldsymbol{Y_{i1}},\dots,\boldsymbol{Y_{ik}},\theta]=\frac{\int_G\int_{\boldsymbol{b_i}}G_i\left[\prod_j f\left(y_{ij1},\dots,y_{ijK}\middle|G_i,\boldsymbol{b_i}\right)\right]g(G_i)h(\boldsymbol{b_i})dG_id\boldsymbol{b_i}}{\int_G\int_{\boldsymbol{b_i}}\left[\prod_j f\left(y_{ij1},\dots,y_{ijK}\middle|G_i,\boldsymbol{b_i}\right)\right]g(G_i)h(\boldsymbol{b_i})dG_id\boldsymbol{b_i}} \tag{5}$$

3. Multivariate Maternal Longitudinal Measure and Cross-Sectional Neonatal Measure

In this mode we consider multivariate longitudinal maternal anthropometry measures along with multiple neonatal anthropometry measures by using a linear mixed model as in the previously presented model and a shared random parameter model as follows:

$$\begin{array}{rcl} Y_{ijk} & = & \beta_{0k}+\beta_{1k}t_{ij}+b_{i0k}+b_{i1k}\left(G_i+t_{ij}\right)+\varepsilon_{ijk} \\ W_{il} & = & \gamma_{0l}+\gamma_{1l}(G_i+t_{i*})+\boldsymbol{b_{ik}}'\boldsymbol{\varphi_l}\ +\epsilon_{il} \end{array} \tag{6}$$

where Y_{ijk} represents the jth measurement for the ith subject of the kth maternal anthropometry measure, with $i=1,\dots,N$, $j=1,\dots,n_i$ and $k=1,\dots,K$. W_{il} represents the lth neonatal anthropometry measure for the ith subject , with $l=1,\dots,L$. $\boldsymbol{b_{ik}}=(b_{i01},bi_{1k},\dots,b_{i0K},b_{i1K})'$ is normally distributed with mean **0** and variance Σ_b, $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, $\epsilon_{il} \sim N(0, \sigma_\epsilon^2)$ and $G_i \sim N(\mu_G, \sigma_G^2)$. t_{i*} is the time between the first visit and delivery, we assume that $t_{i0} = 0$ for all i and that G_i is observed.

The estimation of (6) is done using a two stage approach.

Step 1: Estimate the longitudinal maternal anthropometric portion of the model, by either estimating the multivariate model or using the pairwise approach [34].

Step 2: Fit the neonatal anthropometry model

$$W_{il} = \gamma_{0l} + \gamma_{1l}(G_i + t_{i*}) + \widehat{\boldsymbol{b}}_{\boldsymbol{ik}}{}' \boldsymbol{\varphi}_{\boldsymbol{l}} + \epsilon_{il}$$

via linear regression, using $\widehat{\boldsymbol{b}}_{\boldsymbol{ik}}$, the estimated random effects obtained in Stage I.

Estimations of μ_G and σ_G^2 are obtained from (2).

To predict gestational age we propose $E[G_i | \boldsymbol{Y}_{\boldsymbol{i1}}, \dots, \boldsymbol{Y}_{\boldsymbol{ik}}, \boldsymbol{W}_{\boldsymbol{i}}, \theta]$, where $\boldsymbol{Y}_{\boldsymbol{ik}} = \left(Y_{i1k}, \dots, y_{in_ik}\right)'$, $\boldsymbol{W}_{\boldsymbol{i}} = (W_{i1}, \dots, W_{iL})'$ and $\theta = \{\boldsymbol{\beta}_{\boldsymbol{1}}, \dots, \boldsymbol{\beta}_{\boldsymbol{k}}, \boldsymbol{\gamma}, \boldsymbol{\varphi}_{\boldsymbol{l}}, \Sigma_b, \sigma_\varepsilon^2, \sigma_\epsilon^2, \mu_G, \sigma_G^2\}$, with $\boldsymbol{\beta}_{\boldsymbol{k}} = (\beta_{0k}, \beta_{1k})'$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$.

$$E[G_i | \boldsymbol{Y}_{\boldsymbol{i1}}, \dots, \boldsymbol{Y}_{\boldsymbol{ik}}, \boldsymbol{W}_{\boldsymbol{i}}, \theta] = \frac{\int_G \int_{\boldsymbol{b}_{\boldsymbol{i}}} G_i \left[\prod_j f\left(y_{ij1}, \dots, y_{ijK}, \boldsymbol{w}_{\boldsymbol{i}} \middle| G_i, \boldsymbol{b}_{\boldsymbol{i}}\right)\right] g(G_i) h(\boldsymbol{b}_{\boldsymbol{i}}) dG_i d\boldsymbol{b}_{\boldsymbol{i}}}{\int_G \int_{\boldsymbol{b}_{\boldsymbol{i}}} \left[\prod_j f\left(y_{ij1}, \dots, y_{ijK}, \boldsymbol{w}_{\boldsymbol{i}} \middle| G_i, \boldsymbol{b}_{\boldsymbol{i}}\right)\right] g(G_i) h(\boldsymbol{b}_{\boldsymbol{i}}) dG_i d\boldsymbol{b}_{\boldsymbol{i}}} \qquad (7)$$

Appendix S2

Instructions:

1. The user should replace the text in red below to the data specific to their patient of interest:
    a. Fundal: Gestational Age in weeks for when fundal height was measured
    b. Dates: The dates that correspond to each of the gestational age measurements
    c. Del Date: The delivery date
    d. ***NOTE***: Do not change the format for dates (e.g. MM/DD/YY) or gestational age (XX.XX)
2. Copy all code from **Code Begins** through **Input data below** into R console
3. The final output will be the estimated gestational age

See the example below for details.

<u>GADelCalc Documentation</u>

Description: The function GADelCalc computes the gestational age at delivery from maternal fundal height measurements.

fundal	numeric vector that contains the fundal height measurements in cm.
dates	character vector that contains the measurement dates in the format "%m/%d/%y".
del.date	value of the date of delivery in the format "%m/%d/%y".

Details: A value is returned, which corresponds to the gestational age in weeks of the newborn at delivery.

```
GADelCal<-function(fundal,dates,del.date){
 library(mvtnorm)
 Beta=c(0.9824,0.9626)
 SigmaV=matrix(c(4.8669,-0.1551,-0.1551,0.0060),nrow=2)
 Sigmae=1.7212
 mmGA=19.5275
 vvGA=6.2366
 if(length(fundal)!=length(dates)){stop("The number of measurements and dates do not match  ")}
 dat=as.Date(dates,format = "%m/%d/%y")
 if(length(which(is.na(dat)==T))){stop("The visit dates are in the wrong format, use %m/%d/%y")}
 n.o=length(fundal)
 tt=as.numeric(difftime(dat,dat[1],units='weeks'))
 if(length(tt)<2){stop("Insufficient measurements, please enter at least 2")}
 de.t=as.numeric(difftime(as.Date(del.date,format = "%m/%d/%y"),dat[1],units='weeks'))
 if(is.na(de.t)==T){stop("The delivery date is in the wrong format, use %m/%d/%y")}
```

```
  if(as.Date(del.date,format = "%m/%d/%y")<max(dat)){stop("The delivery date is smaller than a visit
date")}
  Num=0
  Den=0
  for (i in 1:1000){ #Approximating the integral
    GA=i/25
    Z=c(1,GA)
    for (j in 1:(length(tt)-1)){Z=rbind(Z,c(1,GA+tt[j+1]))}
    mm=Z%*%Beta
    vv=Z%*%SigmaV%*%t(Z)+Sigmae*diag(length(tt))
    Num=Num+GA*dmvnorm(fundal,mm,vv)*pnorm(GA,mmGA,vvGA)
    Den=Den+dmvnorm(fundal,mm,vv)*pnorm(GA,mmGA,vvGA)
  }
  GAi=Num/Den
  GaDel=GAi+de.t

  return(GaDel)
}

#************************************
#  Example
#************************************
fundal=c(21.00,24.55,33.00,30.00,34.00,37.00,37.00,37.40,41.00,38.00)
dates=c('7/4/14','7/19/14','8/22/14','9/5/14','9/28/14',
       '9/29/14','10/17/14','10/26/14','10/27/14','11/2/14')
del.date=c('11/5/14')

GADelCal(fundal,dates,del.date)
```